RR 91·03

POINTS OF VIEW ANALYSIS REVISITED: FITTING MULTIDIMENSIONAL STRUCTURES TO OPTIMAL DISTANCE COMPONENTS WITH CLUSTER RESTRICTIONS ON THE VARIABLES

Jacqueline J. Meulman

Peter Verboon

Department of Data Theory

University of Leiden

POINTS OF VIEW ANALYSIS REVISITED: FITTING MULTIDIMENSIONAL STRUCTURES TO OPTIMAL DISTANCE COMPONENTS WITH CLUSTER RESTRICTIONS ON THE VARIABLES

Abstract

Points of view analysis (PVA), proposed by Tucker and Messick in 1963, was one of the first methods to deal explicitly with individual differences in multidimensional scaling, but at some point was apparently superceded by the Carroll and Chang INDSCAL model. This paper argues that a points of view analysis deserves new attention, especially as a technique to analyze group differences, and a streamlined version of the Tucker and Messick procedure is proposed. The separate components in the original procedure are integrated into a comprehensive least squares loss function that is minimized through a convergent algorithm. In addition, two types of nonlinear transformations are incorporated, either with respect to dissimilarities, or for variables from which dissimilarities are derived. Various applications are discussed, where the two types of transformation can be mixed in the same analysis; a quadratic assignment framework is used to evaluate the results.

Key words: points of view, individual differences, group differences, nonlinear multivariate analysis, nonmetric multidimensional scaling, distance components, composite dissimilarities, variable clustering.

1. Introduction

Points of view analysis as proposed by Tucker and Messick (1963) was one of the first methods that sought a compromise between two approaches to the multidimensional scaling (MDS) analysis of dissimilarity data for objects or stimuli obtained from different individuals (sources). One approach was based on group averages, where groups were chosen a priori; the other was an analysis on an individual level. The objective of a points of view analysis is to obtain different multidimensional spaces for different groups of individuals, each having a particular viewpoint about the object interrelationships. Before summarizing the Tucker and Messick approach, and the issues that were raised with respect to the procedure, some preliminary notation must be given.

Dissimilarity data are available for N objects, obtained from $M > 1$ different sources or individuals, $m = 1,\ldots,M$. The dissimilarity data may be represented in two different forms; the first is a symmetric matrix $\mathbf{\Delta}_m = \{\delta_{ijm}\}$, of order NxN, containing the dissimilarities δ_{ijm} between pairs of objects $\{i,j\}$ according to source m. We assume $\mathbf{\Delta}_m$ is normalized so that the sum of squares of the elements is $2N$. The alternative representation is a vector $\boldsymbol{\delta}_m$, of order l, where $l = N(N-1)/2$, containing the dissimilarities d_{ijm} for each object pair for which $i < j$; here, the sum of squares of the elements is set equal to N.

The index set $\{1,\ldots,M\}$ is assumed to be partitioned into subsets J_s, $s = 1,\ldots,r$, giving the sources that form the sth point of view; M_s indicates the number of indices in J_s. For each of the r viewpoints we wish to find a representation space $\mathbf{X}_s$; the dimensionality of $\mathbf{X}_s$ is assumed to be p_s, and the rows of $\mathbf{X}_s$ give the coordinates for the N objects in the representation space for viewpoint s.

A squared distance between a pair of objects $\{i,j\}$ in $\mathbf{X}_s$ is defined by

$$d^2_{ij}(\mathbf{X}_s) = (\mathbf{e}_i - \mathbf{e}_j)'\mathbf{X}_s\mathbf{X}_s'(\mathbf{e}_i - \mathbf{e}_j),$$

where $\mathbf{e}_i$ is the ith column of the NxN identity matrix $\mathbf{I}$. Applying the squared distance function, $D^2(\cdot)$, to map coordinates, $\mathbf{X}_s$, into squared distances gives the matrix formulation:

$$D^2(\mathbf{X}_s) = \boldsymbol{\alpha}_s\mathbf{1}' + \mathbf{1}\boldsymbol{\alpha}_s' - 2\mathbf{X}_s\mathbf{X}_s',$$

with $\mathbf{1}$ an Nx1 vector of all 1's and $\boldsymbol{\alpha}_s$ an Nx1 vector containing the diagonal elements of $\mathbf{X}_s\mathbf{X}_s'$.

Given these preliminaries, the following steps can be distinguished in the original Tucker and Messick (1963) procedure. The dissimilarities are represented in the vectors $\boldsymbol{\delta}_m$, and are regarded as variables that can be subjected to a principal components analysis (PCA) to give principal component scores in r dimensions, $1 \leq r \leq M$, where r denotes the assumed number of different points of view. Each principal component gives a weighted average (linear combination) of the original M dissimilarity variables. Since the principal axes orientation may not be the most appropriate one for displaying different viewpoints, a rotation to simple structure is sought. Transforming the weights obtained in the PCA accordingly, gives a rotated component, denoted as

$$\boldsymbol{\theta}_s = M^{-1} \sum_{m=1}^{M} \mathrm{a}_{ms} \boldsymbol{\delta}_m,$$

$s = 1,\ldots,r$, where a_{ms} is an element of the transformed weight matrix $\mathbf{A} = \{\mathrm{a}_{ms}\}$ that represents simple structure. Next, the weights in $\mathbf{A}$ are used to obtain

$$\boldsymbol{\Theta}_s = M^{-1} \sum_{m=1}^{M} \mathrm{a}_{ms} \boldsymbol{\Delta}_m,$$

$s = 1,\ldots,r$, where each $\boldsymbol{\Theta}_s$ is a differently weighted average (linear combination) of the matrices $\boldsymbol{\Delta}_m$. The final step proposed in Tucker and Messick (1963) consisted of r separate multidimensional scaling analyses fitting Euclidean distances $\mathrm{D}(\mathbf{X}_s)$ to each of the $\boldsymbol{\Theta}_s$.

Ross (1966) criticized the method; among other things, he focuses on the possibility that a point of view might not be a linear combination of judgements of subjects. Cliff (1968), who reviews the method favorably, argues that this criticism can be refuted by realizing that Ross misinterprets the intention of Tucker and Messick with respect to what a point of view really is: it is not a way of looking at the objects, but it is a structure for the objects obtained by an MDS. Another objection by Ross concerns the fact that an arbitrary linear combination might give negative weights for the dissimilarity sources, and could result in a dissimilarity matrix for which no Euclidean solution exists.

Carroll and Chang (1970) proposed the INDSCAL model as an alternative to points of view analysis; the INDSCAL model does not fit weights to the dissimilarity sources, but fits weights for the dimensions of an unknown space **X** common to all sources. Carroll and Chang state that PVA is little more powerful than doing separate scalings, and question the fact that no explicit assumptions are made about the possible communality of the multidimensional structures. Since its introduction, the INDSCAL model seems to have become the dominant model to analyze individual differences. Yet, it is the purpose of the present paper to show there is still room for the PVA model, explicitly when one is interested in finding subsets of individuals (clusters of sources) that have the same frame of reference. PVA is truly different from doing separate scalings since we do not wish to assign the sources to points of view *a priori*; instead, we will perform a clustering task that assigns the sources to different points of view. With respect to the issue of weights possibly being negative, it is guaranteed in the procedure described below that the weights are always positive; moreover, they turn out to be Tucker's (1951) congruence coefficients between each separate dissimilarity source and the distances fitted in the point of view.

In the INDSCAL model, individual differences are defined on dimensions of the common space; points of view are defined on the distances, and the spaces can, but need not, be interpreted in terms of dimensions. It is also possible to look at clusters, structures, or more and other directions than the principal dimensions. The concept of separate point of view spaces will be considered as an alternative for the INDSCAL common space. When the data are truly high-dimensional, the INDSCAL model might require many dimensions in the common space, and in that case, a representation in a number of different points of view, each having only two or three dimensions, might be much easier to interpret visually. In the INDSCAL model the dimensions are not assumed to be uncorrelated; in points of view analysis the distances in the fitted structures are not assumed to be uncorrelated. Whether multiple points of view are really different or actually very similar can be investigated by using quadratic assignment procedures, as discussed, for example, in Hubert (1987).

2. A Comprehensive Objective Function for Points of View Analysis

In the previous section it was shown that points of view analysis deals with three different tasks: the first is to find principal components and weights applying the PCA model to given dissimilarity variables; the second is to find an optimal rotation to simple structure, and the final task is to find optimal representation spaces for the objects on the basis of the rotated components (the composite dissimilarities). In this section a least squares loss function will be introduced that integrates these different optimization tasks. The loss function is defined on the distances in the representation spaces, and thus fits into the STRESS framework, for which Kruskal (1964a,b) has laid the foundation.

If $\|\cdot\|^2$ denotes a least squares discrepancy measure such that

$$\|a_{ms}\mathbf{\Delta}_m^* - D(\mathbf{X}_s)\|^2 = \mathrm{tr}\,(a_{ms}\mathbf{\Delta}_m^* - D(\mathbf{X}_s))'(a_{ms}\mathbf{\Delta}_m^* - D(\mathbf{X}_s)),$$

the PVA loss function can be written as

$$\mathrm{STRESS}(\mathbf{A};\mathbf{\Delta}_1^*,\ldots,\mathbf{\Delta}_M^*;\mathbf{X}_1,\ldots,\mathbf{X}_r) = M^{-1}\sum_{s=1}^{r}\sum_{m\in J_s}\|a_{ms}\mathbf{\Delta}_m^* - D(\mathbf{X}_s)\|^2, \tag{1}$$

which is a function of three sets of parameters. For the moment, consider $\mathbf{\Delta}_1^*,\ldots,\mathbf{\Delta}_M^*$ as given, so without loss of generality $\mathbf{\Delta}_m^* = \mathbf{\Delta}_m$; then the loss function has to be minimized over the weight matrix $\mathbf{A}$ and the points of view $\mathbf{X}_1,\ldots,\mathbf{X}_r$. Since a perfect, but trivial, solution is easily obtained by setting $\mathbf{A} = \mathbf{0}$ and $\mathbf{X}_s = \mathbf{0}$, we require, without loss of generality, that $\mathrm{tr}\,(\mathbf{X}_s'\mathbf{X}_s) = 1$. The loss function must also be minimized over $\mathbf{A} = \{a_{ms}\} \in \Omega$, where Ω is the set of all *restricted* weight matrices that give some form of simple structure.

The objective function in (1) will be minimized by constructing a convergent algorithm using various components from the majorization approach to multidimensional scaling (De Leeuw & Heiser, 1980; Meulman, 1986; De Leeuw, 1988). In the following, the components will be discussed.

Fitting the Multidimensional Structures

The overall minimization problem in (1) can be partitioned into several parts. First of all, for fixed A and $\mathbf{\Delta}_1^*,...,\mathbf{\Delta}_M^*$, the problem of finding $\mathbf{X}_1,...,\mathbf{X}_r$ consists of r separate MDS tasks. For each point of view we have to minimize

$$\mathrm{STRESS}(\mathbf{X}_s) = M^{-1} \sum_{m \in \mathrm{J}_s} \|\mathrm{a}_{ms}\mathbf{\Delta}_m^* - \mathrm{D}(\mathbf{X}_s)\|^2, \tag{2}$$

which is a function of $\mathbf{X}_s$ only. The objective function (2) can be simplified in the following way. Define the composite dissimilarity matrix for the sources that constitute the sth point of view as

$$\mathbf{\Theta}_s = M_s^{-1} \sum_{m \in \mathrm{J}_s} \mathrm{a}_{ms}\mathbf{\Delta}_m^*;$$

then (2) can be written as

$$\mathrm{STRESS}(\mathbf{X}_s) = M^{-1} \Big[\Big(\sum_{m \in \mathrm{J}_s} \|\mathrm{a}_{ms}\mathbf{\Delta}_m^* - \mathbf{\Theta}_s\|^2\Big) + M_s \|\mathbf{\Theta}_s - \mathrm{D}(\mathbf{X}_s)\|^2\Big], \tag{3}$$

where M_s indicates the number of sources assigned to viewpoint s. The first term on the right-hand side of (3) gives stress due to heterogeneity of sources within point of view s; the second term gives the group stress, with respect to the optimally aggregated dissimilarity matrix $\mathbf{\Theta}_s$.

Because the loss due to heterogeneity is a constant term with respect to $\mathbf{X}_s$, (3) is minimized by minimizing $\|\mathbf{\Theta}_s - \mathrm{D}(\mathbf{X}_s)\|^2$ over $\mathbf{X}_s$. The latter can be done by using the majorization algorithm for MDS in its simplest form (e.g., see De Leeuw, 1988). For each representation space $\mathbf{X}_s$ we compute, from a starting point $\mathbf{X}_s^{\mathrm{o}}$, the Guttman transform $\tilde{\mathbf{X}}_s$:

$$\tilde{\mathbf{X}}_s = n^{-1}\mathrm{B}(\mathbf{X}_s^{\mathrm{o}})\mathbf{X}_s^{\mathrm{o}}. \tag{4}$$

The elements of the NxN matrix $\mathrm{B}(\mathbf{X}_s^{\mathrm{o}})$ can be defined in terms of the elements of two auxiliary matrices: the NxN matrix, $\mathrm{B}^{\mathrm{o}}(\mathbf{X}_s^{\mathrm{o}})$, whose elements are

$$\mathrm{b}_{ij}^{\mathrm{o}}(\mathbf{X}_s^{\mathrm{o}}) = M_s^{-1} \sum_{m \in \mathrm{J}_s} \mathrm{a}_{ms}\delta_{ijm}^*/\mathrm{d}_{ij}(\mathbf{X}_s), \quad \text{if } i \neq j \text{ and } \mathrm{d}_{ij}(\mathbf{X}_s) \neq 0,$$
$$\mathrm{b}_{ij}^{\mathrm{o}}(\mathbf{X}_s^{\mathrm{o}}) = 0 \quad \text{if } \mathrm{d}_{ij}(\mathbf{X}_s^{\mathrm{o}}) = 0,$$

and the NxN diagonal matrix $\mathrm{B}^*(\mathbf{X}_s^{\mathrm{o}})$, with diagonal elements

$$b^*_{ii}(\mathbf{X}^o_s) = \mathbf{1}'\mathrm{B}^o(\mathbf{X}^o_s)\mathbf{e}_i.$$

Then,

$$\mathrm{B}(\mathbf{X}^o_s) = \mathrm{B}^*(\mathbf{X}^o_s) - \mathrm{B}^o(\mathbf{X}^o_s).$$

The theory of the majorization algorithm for MDS guarantees that

$$\|\mathbf{\Theta}_s - \mathrm{D}(\tilde{\mathbf{X}}_s)\|^2 \le \|\mathbf{\Theta}_s - \mathrm{D}(\mathbf{X}^o_s)\|^2,$$

so by repeatedly computing the Guttman transform, and setting $\mathbf{X}^o_s = \tilde{\mathbf{X}}_s$ for each new update, a series of convergent configurations is obtained, until STRESS($\mathbf{X}^o_s$)–STRESS($\tilde{\mathbf{X}}_s$) $\le \varepsilon$, with ε some preset small value, and the (possibly local) minimum of (3) is achieved.

Because the minimization of (3) is only a part of the overall problem (1), there is no need to converge to the minimum in each step of the algorithm that updates the representation spaces; for each representation space, a single Guttman transform $\tilde{\mathbf{X}}_s$ suffices to decrease the loss in (1), and when STRESS($\mathrm{A};\mathbf{\Delta}^*_1,\ldots,\mathbf{\Delta}^*_M;\mathrm{X}_1,\ldots,\mathrm{X}_r$) has attained its minimum with respect to the preset small value ε, so must STRESS($\mathbf{X}_s$).

Assigning the Sources to Different Points of View

The second step in the algorithmic scheme should minimize (1) over $\mathrm{A} \in \Omega$, for fixed $\mathbf{\Delta}^*_1,\ldots,\mathbf{\Delta}^*_M$ and $\mathrm{X}_1,\ldots,\mathrm{X}_r$. In this paper, we explicitly require that the index sets J_s are mutually exclusive, so that each dissimilarity source contributes to only one point of view, but other approaches are also possible. To solve for A, we first construct $\tilde{\mathrm{A}}$, minimizing

$$\mathrm{STRESS}(\mathrm{A}) = M^{-1}\sum_{s=1}^{r}\sum_{m=1}^{M} \|(a_{ms}\mathbf{\Delta}^*_m - \mathrm{D}(\mathbf{X}_s)\|^2,$$

over A unrestricted. By setting the partial derivatives with respect to a_{ms} equal to zero, we obtain

$$\tilde{a}_{ms} = (2N)^{-1}\,\mathrm{tr}\,(\mathbf{\Delta}^*_m\,\mathrm{D}(\mathbf{X}_s)). \tag{5}$$

(The dissimilarity matrices were assumed to be normalized so that the sum of squares of the elements is $2N$.) The estimates in (5) are positive by definition, and because $\mathrm{tr}\,(\mathbf{X}_s'\mathbf{X}_s) = 1$, the sum of squares of the elements in $D(\mathbf{X}_s)$ is equal to $2N$, so that (5) gives Tucker's (1951) congruence coefficient between the individual dissimilarities and the distances in the sth point of view.

When each source is assigned to only one point of view, the index set $\{1,...,M\}$ must be partitioned into *non-overlapping* subsets; in that case the constraints on the weight matrix $\mathbf{A}$ can be written in the form $\mathbf{A} = \mathbf{WG}$, where $\mathbf{W}$ is a diagonal matrix, of order M x M, containing a single weight w_{mm} for each source $\mathbf{\Delta}_m^*$ on its diagonal, and $\mathbf{G}$ is a binary and orthogonal matrix, of order M x r, that assigns each source $\mathbf{\Delta}_m^*$ to one of the r viewpoints. So we minimize

$$\|\tilde{\mathbf{A}} - \mathbf{WG}\|^2,$$

over $\mathbf{W}$ and $\mathbf{G}$. This function, which finds non-overlapping clusters of dissimilarity sources $\mathbf{\Delta}_m^*$, can be fitted row after row. The diagonal elements of $\mathbf{W}$ are found as $w_{mm} = \max(\tilde{a}_{m1},...,\tilde{a}_{mr})$; in the (unlikely) case that some values in the mth row of $\mathbf{A}$ are exactly equal, we would need an "untie" procedure. Next, $\mathbf{G}$ is obtained by setting $\mathbf{G}_{ms} = 1$ if $\tilde{a}_{ms} = w_{mm}$, and $\mathbf{G}_{ms} = 0$ otherwise, and the restricted weight matrix is set to $\mathbf{A} = \mathbf{WG}$.

3. Nonlinear Generalizations

In the previous section it was described how the general loss function (1) is minimized over $\mathbf{X}_1,...,\mathbf{X}_r$ and $\mathbf{A}$; in this section nonlinear generalizations (i.e., finding the optimal $\mathbf{\Delta}_m^*$) will be discussed. There are two different approaches. First, the relation with Gifi's (1990) approach to nonlinear principal components analysis will be considered; it will be shown that this approach, when applied to distance variables, reduces to nonmetric scaling. Next, a second form of transformation will be proposed, originating from the distance approach to nonlinear multivariate analysis (Meulman, 1986). Finally, the two possibilities will be combined.

In Gifi (1990) a system of nonlinear MVA techniques is developed that has the notion of homogeneity as starting point. In a principal components analysis, the NxM data matrix $\mathbf{Z}$ is analyzed, whose columns are defined by Nx1 vectors $\mathbf{z}_m$, m=1,...,M, that contain observations on the variables assumed to have means of zero and sum of squares of one. The measurements on the objects for the M variables define the rows in $\mathbf{Z}$. In the Gifi system, a nonlinear PCA in r dimensions can be written in the form of the loss function:

$$\text{STRIFE}(\mathbf{q}_1,\ldots,\mathbf{q}_M;\mathbf{x}_1,\ldots,\mathbf{x}_r;\mathbf{A}) = M^{-1} \sum_{s=1}^{r} \sum_{m=1}^{M} \|\mathrm{a}_{ms}\mathbf{q}_m - \mathbf{x}_s\|^2, \tag{6}$$

that has to be minimized over $\mathbf{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_r\}$, constrained so that $\mathbf{X'X} = \mathbf{I}$, over $\mathbf{A} = \{\mathrm{a}_{ms}\}$, and over $\mathbf{q}_1,\ldots,\mathbf{q}_M$, satisfying $\mathbf{q}'_m\mathbf{q}_m = 1$ and $\mathbf{q}_m \in \Gamma_m$, where Γ_m indicates the set of admissible transformations of the given variable $\mathbf{z}_m$.

The class of transformations may be defined differently for each variable $\mathbf{z}_m$, and includes nominal transformations (that preserve equal values in $\mathbf{z}_m$ by giving ties in $\mathbf{q}_m$), monotonic transformations (that maintain the order of the elements of $\mathbf{z}_m$ in $\mathbf{q}_m$), and linear transformations (which implies setting $\mathbf{q}_m = \mathbf{z}_m$, since it was required that $\mathbf{q}'_m\mathbf{q}_m = 1$). The loss function says that each weighted transformed variable $\mathrm{a}_{ms}\mathbf{q}_m$ should resemble the unknown $\mathbf{x}_s$ as closely as possible, where $\mathbf{x}_s$ turns out to be the normalized sth principal component (Gifi, 1990, ch. 3)

Transformation of the variables in PCA can be applied straightforwardly to points of view analysis, when the latter is regarded as a components analysis of dissimilarity variables. When $\boldsymbol{\delta}^*_m$ denotes the optimal transformation of a given dissimilarity variable $\boldsymbol{\delta}_m$, then (6) could be viewed as a nonlinear variety of the original PVA procedure (Verboon & Van der Kloot, 1989), replacing $\mathbf{q}_m$ by $\boldsymbol{\delta}^*_m$ and $\mathbf{x}_s$ by $\boldsymbol{\theta}_s$. The weight matrix $\mathbf{A}$ will in general not give a simple structure, but as is remarked in Gifi (1990, ch. 4; ch. 10), it is possible to generalize (6) to restricted $\{\mathrm{a}_{ms}\}$, requiring, for instance, some a_{ms} to be zero.

When the variables are dissimilarities, the components $\mathbf{x}_s$ are optimal with respect to replacing the M variables $\boldsymbol{\delta}^*_m$ by a fewer number of (latent) dissimilarity variables. The particular linear combination, however, will in general not be optimal for the final step in a PVA (i.e., obtaining a low-dimensional space in which the distances between the objects resemble the composite dissimilarities in $\mathbf{x}_s$ as closely as possible). Therefore, we propose to find optimal distance components, minimizing

$$\mathrm{STRESS}(\boldsymbol{\delta}_1^*,\ldots,\boldsymbol{\delta}_M^*;\mathbf{X}_1,\ldots,\mathbf{X}_r;\mathbf{A}) = M^{-1}\sum_{s=1}^{r}\sum_{m\in \mathrm{J}_s} \|\mathrm{a}_{ms}\boldsymbol{\delta}_m^* - \mathbf{d}(\mathbf{X}_s)\|^2, \tag{7}$$

satisfying $\boldsymbol{\delta}_m^{*\prime}\boldsymbol{\delta}_m^* = N$ and $\boldsymbol{\delta}_m^* \in \Lambda_m$. Here, Λ_m denotes the set of admissible transformations of $\boldsymbol{\delta}_m$; when the variables are dissimilarities, Λ_m is typically chosen as a set of monotonic transformations. The vector $\mathbf{d}(\mathbf{X}_s) = \{\mathrm{d}_{ij}(\mathbf{X}_s) \text{ for } \mathrm{i} < \mathrm{j}\}$ contains the lower diagonal elements of the matrix $\mathrm{D}(\mathbf{X}_s)$ in some predetermined order, so (7) is a special case of the general loss function (1).

Including general monotonic transformations of the dissimilarities has brought us to the domain of nonmetric multidimensional scaling, originating from Shepard (1962a; 1962b) and Kruskal (1964a; 1964b). In fact, the KYST program (Kruskal, Young, & Seery, 1973) can be used to fit a single point of view; (7) creates the possibility of multiple points of view by applying cluster restrictions to **A**.

The transformed dissimilarities are called pseudo-distances, and are usually restricted to be monotonic with the given vector $\boldsymbol{\delta}_m$. When $\tilde{\boldsymbol{\delta}}_m$ denotes the unrestricted estimate, obtained by setting partial derivatives with respect to $\boldsymbol{\delta}_m^*$ in (7) equal to zero,

$$\tilde{\boldsymbol{\delta}}_m = \mathbf{d}(\mathbf{X}_s) / \mathrm{a}_{ms} \text{ if } m \in \mathrm{J}_s.$$

Next we minimize

$$\|\tilde{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_m^*\|^2,$$

satisfying $\boldsymbol{\delta}_m^{*\prime}\boldsymbol{\delta}_m^* = N$, over $\mathbf{d}_m^* \in \Lambda_m$, where Λ_m denotes the set of admissible monotonic transformations of $\boldsymbol{\delta}_m$; Λ_m can be chosen as the set of general monotonic transformations as in Kruskal (1964a,b), but another possibility is to choose Λ_m as the set of monotonic spline transformations of a particular degree, with a prechosen number of knots, as in Ramsay (1982a,b).

The second nonlinear generalization is of a different nature. Going back to (6) in the analysis of the data matrix $\mathbf{Z}$, instead of approximating $a_{ms}\mathbf{q}_m$ directly, we approximate $D(a_{ms}\mathbf{q}_m)$, the distances that a weighted variable generates. This approach is in line with Meulman (1986), and applied to PVA it implies the minimization of

$$\mathrm{STRESS}(\mathbf{q}_1,\ldots,\mathbf{q}_M;\mathbf{X}_1,\ldots,\mathbf{X}_r;\mathrm{A}) = M^{-1}\sum_{s=1}^{r}\sum_{m\in \mathrm{J}_s} \|D(a_{ms}\mathbf{q}_m) - D(\mathbf{X}_s)\|^2. \tag{8}$$

Due to the homogeneity of the Euclidean distance function, $D(a_{ms}\mathbf{q}_m) = a_{ms}D(\mathbf{q}_m)$, and by setting $\mathbf{\Delta}^*_m = D(\mathbf{q}_m)$, (8) turns into another special case of (1).

The optimal transformations of the variables are obtained by the following procedure, derived from the majorization algorithm for MDS with restrictions on the configuration (De Leeuw & Heiser, 1980). In the minimization of (8), the representation spaces $\mathbf{X}_s$ generate target values $d_{ij}(\mathbf{X}_s)$ that have to be approximated, and $\mathbf{q}_m$ is considered a restricted one-dimensional configuration. When $\mathbf{q}^o_m$ denotes a starting point that satisfies the constraints, the unrestricted estimate $\tilde{\mathbf{q}}_m$ is obtained by computing the so-called reversed Guttman transform (Meulman, 1986) defined by analogy with (4) as

$$\tilde{\mathbf{q}}_m = n^{-1}\mathrm{B}(\mathbf{q}^o_m)\mathbf{q}^o_m,$$

where

$$\mathrm{B}(\mathbf{q}^o_m) = \mathrm{B}^*(\mathbf{q}^o_m) - \mathrm{B}^o(\mathbf{q}^o_m).$$

Here the elements of $\mathrm{B}^o(\mathbf{q}^o_m)$ are given by

$$b^o_{ij}(\mathbf{q}^o_m) = d_{ij}(\mathbf{X}_s) \,/\, d_{ij}(\mathbf{q}^o_m) \qquad \text{if } m \in \mathrm{J}_s,\ i\neq j \text{ and } d_{ij}(\mathbf{q}^o_m) \neq 0;$$

$$b^o_{ij}(\mathbf{q}^o_m) = 0 \qquad \text{if } d_{ij}(\mathbf{q}^o_m) = 0,$$

and the diagonal elements of $\mathrm{B}^*(\mathbf{q}^o_m)$ as

$$b^*_{ii}(\mathbf{q}^o_m) = \mathbf{1}'\mathrm{B}^o(\mathbf{q}^o_m)\mathbf{e}_i.$$

The basic theory of the majorization algorithm says that

$$\|a_{ms}D(\tilde{q}_m) - D(\mathbf{X}_s)\|^2 \leq \|a_{ms}D(q_m^0) - D(\mathbf{X}_s)\|^2.$$

Using the results from De Leeuw and Heiser (1980), it can be shown that also

$$\|a_{ms}D(\hat{q}_m) - D(\mathbf{X}_s)\|^2 \leq \|a_{ms}D(q_m^0) - D(\mathbf{X}_s)\|^2,$$

when

$$\hat{q}_m = \operatorname{argmin} \|\tilde{q}_m - \mathbf{q}_m\|^2,$$

where the minimization is over $\mathbf{q}_m \in \Gamma_m$, satisfying $\mathbf{q}_m'\mathbf{q}_m = 1$, with Γ_m the set of all admissible transformations of a given variable $\mathbf{z}_m$. As in (6) the transformations may be nominal (preserving ties), monotonic (preserving order), or linear (setting $\mathbf{q}_m = \mathbf{z}_m$). The normalization $\mathbf{q}_m'\mathbf{q}_m = 1$ is equivalent to $\delta_m^{*\prime}\delta_m^* = N$ and $\Sigma_i\Sigma_j\, \delta_{ijm}^{*2} = 2N$.

In this application of PVA, the sources $\mathbf{\Delta}_m^*$ are generated by the variables, the columns of the transformed multivariate data matrix $\mathbf{Q} = \{\mathbf{q}_1,...,\mathbf{q}_M\}$, and different subsets of variables are assumed to generate different viewpoints about the interrelationships between the objects, the rows of $\mathbf{Q}$. The weights in $\mathbf{A}$ could be viewed as a replacement for the squared component loadings in an ordinary principal components analysis, and their means as equivalent to the eigenvalues.

Combining (7), where dissimilarities are transformed, and (8), where variables are transformed, in the form of the general objective function (1), creates the possibility of analyzing mixtures of data that consist of dissimilarities and multivariate data for the same set of N objects. If a multivariate data matrix $\mathbf{Z}$ is available at the outset, we still have the choice of considering each $\mathbf{\Delta}_m^*$ either as a monotonic transformation of $D(\mathbf{z}_m)$, or as $D(\mathbf{q}_m)$, with either nominal, ordinal, or numerical transformations for $\mathbf{q}_m$.

4. Points of View Analysis in Action

To discuss the properties of PVA as presented in this paper in more detail, data will be analyzed that have been obtained from a questionnaire study among the members of the Second Chamber of the Dutch Parliament in 1979-1980 (the data were kindly made available by the Department of Political Science of the University of Leiden). In this study, 139 of the 150 members of parliament (MP's) participated; they belong to 11 different political parties, and a short description is given in Table 1. From the extensive questionnaire, several variables have been chosen for different applications of PVA, the data always pertaining to the same set of 139 MP's. The parties in Table 1 have been ordered by using averages within parties derived from a variable that gives the position that the MP's assigned themselves to on a political left-right scale, with a range from 1 (extremely left) to 9 (extremely right).

TABLE 1
Political Parties in the Dutch Parliament in 1979-1980,
Party Membership of the Respondents, and Mean Self Rating within Parties
on the Left (1)-Right (9) Scale in the Questionnaire

Label	Party	Description	Nr. of Respondents	Mean Left-Right
	CPN	Communists	0	
2	PSP	Pacifistic socialists	1	1.00
3	PPR	Radicals	3	1.33
4	PvdA	Social democrats	53	2.70
5	DS70	Social democrats (economically conservative)	1	3.00
6	D66	Liberals (economically progressive)	8	3.25
7	ARP	Protestants	12	4.00
8	KVP	Catholics	24	4.21
9	CHU	Protestants	9	5.00
10	VVD	Liberals (economically conservative)	25	5.04
11	GPV	Very conservative Calvinists	1	7.00
12	SGP	Very conservative Calvinists	2	8.50
	BP	Farmers Party	0	
	CDA	Merger of ARP, KVP and CHU		

In the first application, data are analyzed that were expressed in values on so-called sympathy scales, ranging from 0 (extremely unappealing) to 100 (extremely appealing): each MP gave a score to each of the parties residing in Parliament in 1979 (described in Table 1). In the second application, data are used that give the MP's position with respect to 8 political issues, measured by self-ratings on a 9-point scale. The lower and upper end of the scales for the political issues is given in Table 2, as well as the marginal frequencies of the categories.

TABLE 2
Political Issues in the Questionnaire

	Issue	Lower End (1)	(9) Upper End
1	Development aid	The government should spend much more money on development aid	The government should spend much less money on development aid
2	Abortion	The government should prohibit abortion under all circumstances	Every woman has the right to decide for herself on this matter
3	Law & Order	The government takes too rigorous measures against disturbances of the Queen's peace	The government should take even more rigorous measures
4	Income Differences	The differences in income should remain as they are	The differences in income should become much smaller
5	Employees' Participation	Only management should decide in matters that concern the company	Employees should have their say in decisions
6	Taxes	Taxes should be raised so that more money will become available for public provisions	Taxes should be lowered so that everybody can decide for him/herself
7	Armies	The government should insist on reducing armed forces, even if this would imply risk	The government should maintain strong armed forces.
8	Nuclear Energy	The number of nuclear power plants should be increased rapidly	Nuclear power plants should not be built at all

Marginal Frequencies of Categories

Issue	1	2	3	4	5	6	7	8	9
1	29	35	25	20	24	2	3	1	0
2	3	14	6	7	11	7	16	28	47
3	6	11	23	23	46	14	12	3	1
4	1	4	1	9	12	23	31	34	24
5	0	1	0	3	9	10	28	45	43
6	11	11	29	23	25	19	14	5	2
7	16	13	17	21	25	12	16	7	10
8	2	3	10	18	24	12	20	27	23

In the interpretation of the results, the skewness of some of the distributions of the MP's over the categories should be taken into account.

Analysis of the Sympathy Scales

The first application concerns the sympathy data. Here we wish to investigate whether members of parliament being in different positions in the political spectrum possibly have a different system of sympathies towards the other parties. To explore this question, the MP's are considered judges of the interrelationships between the political parties, so we have to consider the MP's as the columns of the data matrix or the variables, and the parties as the rows or the objects. This implies we have 139 dissimilarity matrices, one for each MP, of order 14 x 14, since there are 14 different parties judged (there are 3 more parties than the number of parties for which we have MP's, because the CPN and BP MP's did not participate in the questionnaire, and the CDA only acts as a stimulus party, since it is the result of a merger between the ARP, KVP and CHU). We chose to define $\mathbf{D}_m^*$ as $D(\mathbf{q}_m)$ as in (8) to find a transformation of the variables; although other monotonic transformations of the data in $\mathbf{z}_m$ could have been considered, to obtain smooth transformations, we chose to define Γ_m as the set of monotonic spline transformations (as in the approach to PCA in Winsberg & Ramsay, 1983; Ramsay, 1989). Second-degree splines with one interior knot were used, which fixes the number of parameters fitted for each sympathy scale to be equal to three.

Two points of view were considered; a single point of view clearly did not fit the data, and two points of view were considered sufficient in terms of goodness-of-fit. The mean weight (that replaced the average eigenvalue) is 0.901, the group stress in (3) is 0.012 for the first viewpoint and 0.026 for the second viewpoint, and the stress due to heterogeneity within groups is 0.058 and 0.087, respectively. Table 3 gives the distribution of the MP's over the two points of view, and weights and badness-of-fit values that have been averaged over sources that belong to the same political party. The overall stress value, described in (1), is the sum of the partitioned stress values that are given in (3), and is equal to the weighted sum over parties, divided by M, the total number of sources.

TABLE 3
Distribution of MP's over Two Points of View, Mean Weights and Partitioned Loss: Total Stress = Heterogeneity + Group Stress

	First Point of View				Second Point of View			
Source	#	Mean Weight	Mean Stress	Hetero-geneity	#	Mean Weight	Mean Stress	Hetero-geneity
PSP	1	0.886	0.215	0.170	0			
PPR	3	0.930	0.135	0.112	0			
PvdA	52	0.930	0.133	0.113	1	0.948	0.101	0.106
DS70	0				1	0.883	0.221	0.157
D66	7	0.857	0.263	0.199	1	0.846	0.284	0.191
ARP	0				12	0.884	0.216	0.165
KVP	0				24	0.914	0.162	0.135
CHU	0				9	0.927	0.140	0.124
VVD	1	0.843	0.289	0.216	24	0.850	0.275	0.197
GPV	0				1	0.930	0.134	0.118
SGP	0				2	0.806	0.348	0.235

	Total Stress	Hetero-geneity	Group Stress	Mean Weight
First point of view	0.070	0.058	0.012	0.920
Second point of view	0.113	0.087	0.026	0.885
Overall	0.183	0.145	0.038	0.901

The first point of view is defined by 64 sources and seems to represent the sympathy of the MP's who, according to their self ratings, are left from the center in the political spectrum. There are 4 exceptions: one member of the PvdA and one member of D66 are in the second point of view, and so is the MP of DS70 (who, accordingly, might be considered not as left-wing as the self rating suggests). Also, one member of the VVD, whose other 24 MP's are in the second point of view, is in the first point of view. In the configurations for the point of views, the objects are represented by 14 political party points.

The first point of view is displayed in Figure 1; it is made up by members of the PvdA (52 out of 64), and the MP's of the PSP, PPR, and D66 and could be called 'the left-wing point of view'. The object point for the PvdA party is represented at the left-hand side of the Figure: on the average, the largest sympathy in the first point of view is for the PvdA. Next, moving to the right, there are two parties that are separated from each other in the second dimension: the PPR that is more left-wing than the PvdA, and D66 that is more to the center. More distant are the even more left-wing PSP, in the second dimension close to the PPR, and the more center

ARP, in the second dimension close to D66; next follow the CPN and the VVD. Distances become quite large between the PvdA on the one hand, and the denominational parties KVP, CHU, GPV, and SGP, on the other. The CDA is located in between the three participating parties (but closer to the KVP and CHU than to the ARP). There is also little sympathy for DS70, which is understandable since it is a conservative secession from the PvdA, and there is no sympathy at all for the very right-wing BP. The overall configuration can be captured in an elliptical structure (as drawn in Figure 1); starting at the point for the CPN, and moving clockwise along the ellipse in the direction of the PSP, the left-right order is recovered (see Table 1). The VVD does not fit on the ellipse; there is more sympathy for the VVD in the left-wing point of view than can be explained from the left-right scale. From a substantive point of view, the political structure can easily be understood by looking at the distances between the political parties, but the dimensions cannot be given a politically relevant interpretation.

Inspecting the data of the MP of the VVD who fits in the viewpoint, shows that this MP orders the parties almost perfectly reversed compared to the average MP of the PvdA. Compared to the average MP of the VVD, this MP has much less sympathy for PvdA, D66, and ARP, and much more for GPV and SGP. The MP of the PvdA who does not fit into the left-wing point of view, orders the parties with decreasing sympathy as (CHU KVP ARP CDA) (GPV SGP) (PvdA D66 VVD) DS70 PPR (PSP CPN BP). A first conclusion might be that the point of view analysis perfectly discovered a coding error in the party membership variable; however, this is unlikely considering other data available, so perhaps it should be concluded that this MP is changing his or her political affiliation.

Figure 2 shows the second point of view; it could be called the 'center-right-wing point of view'. The majority of the MP's that adhere to this viewpoint belong to the parties that merged into the CDA; MP's of parties that are more conservative also fit this viewpoint. One part of the Christian democrats has great sympathy for parties that are more left-wing (PvdA and D66), while the others have more sympathy to parties that are more to the right (VVD, GPV, SGP, and DS70). There is hardly any sympathy for the small extreme left-wing parties PPR,

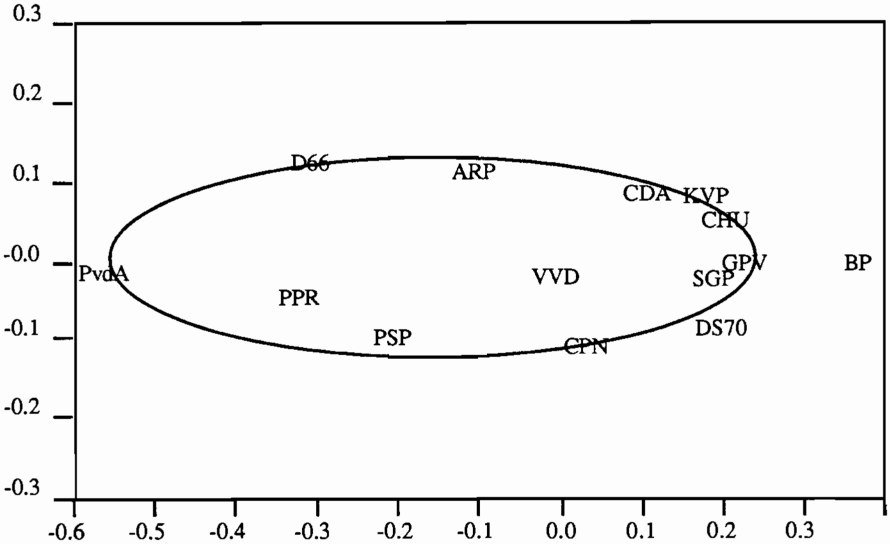

Figure 1.
First viewpoint in analysis of political sympathy scales:
Left-wing point of view.

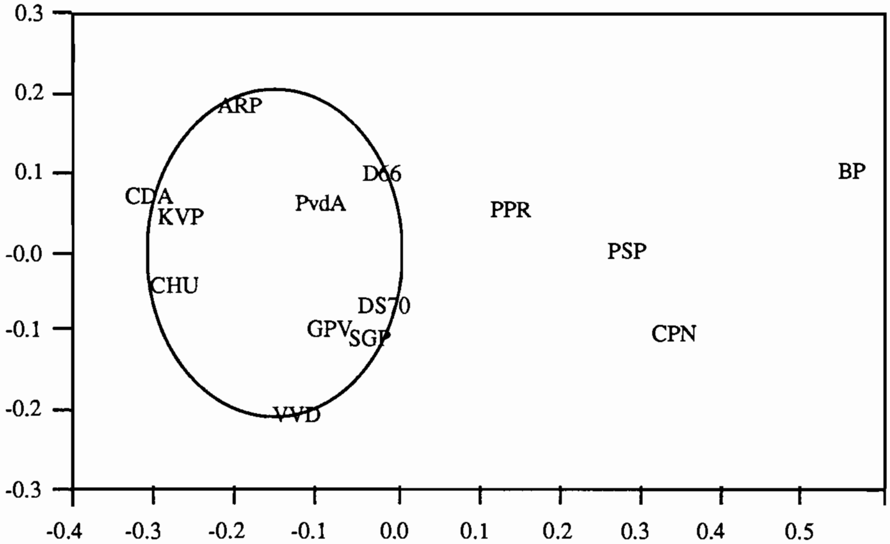

Figure 2.
Second viewpoint in analysis of political sympathy scales:
Center-right-wing point of view.

PSP, and CPN, and the extreme right-wing BP. The latter political parties fall outside the ellipse that orders the parties from left to right when we start at the D66 point and move counter-clockwise towards the ARP; in the center-right-wing point of view there is more sympathy for the PvdA than can be explained from the left-right scale.

TABLE 4
Distances that Differ Most in the Two Points of View

Distance between		First Point of View	Second Point of View
BP	CHU	0.17	0.85
BP	KVP	0.20	0.84
BP	CDA	0.30	0.87
BP	SGP	0.16	0.68
BP	GPV	0.14	0.63
BP	D70	0.18	0.61
CDA	CPN	0.20	0.67
PvdA	D70	0.77	0.14
PvdA	GPV	0.79	0.17
PvdA	SGP	0.77	0.15
PvdA	CHU	0.78	0.20
PvdA	CDA	0.65	0.22
PvdA	KVP	0.75	0.19

The major agreement between the two viewpoints seems to be the antipathy towards the BP. To compare them in more detail, Table 4 gives the distances between parties that differ more than twice the average difference. In the left-wing point of view, the distances between the extremely right-wing BP and CHU, KVP, CDA, SGP, GPV, and DS70 are small; in the center-right-wing point of view they are large. The same is true for the distance between CDA and CPN. It is exactly the other way around for the distances between PvdA and CHU, KVP, CDA, SGP, GPV, and DS70; in the left-wing point of view they are large, and in the center-right-wing point of view they are small.

Analysis of the Political Issues

In the second example a points of view analysis is applied that accommodates the two different types of nonlinear transformations. The first question that comes to mind when analyzing political issues, is whether they can be captured in a single point of view (dominated by the left-right dimension), or do they need a second. To investigate the homogeneity within parties at the same time, a variable indicating party membership was included in the analysis in the following way. First an indicator matrix **B** was constructed, with N rows and 11 columns, indicating for each MP to which of the 11 parties (s)he belongs. From the indicator matrix **B**, dissimilarities between the MP's were derived; although any dissimilarity measure could have been considered, the chi-square distance was selected that is also used in homogeneity analysis (or multiple correspondence analysis). The particular use here is similar to Gifi's (1990) approach to principal components analysis, being a mixture of PCA as in (6) and homogeneity analysis. The squared χ^2- distance between two MP's i and j is defined by

$$\chi^2_{ij}(\mathbf{B}) = (\mathbf{e}_i - \mathbf{e}_j)'\mathbf{B}\mathbf{M}^{-1}\mathbf{B}'(\mathbf{e}_i - \mathbf{e}_j) = \mathrm{d}^2_{ij}(\mathbf{B}\mathbf{M}^{-1/2}), \tag{9}$$

where the matrix $\mathbf{M}^{-1} = (\mathbf{B}'\mathbf{B})^{-1}$ is a diagonal matrix that has the inverse of the column marginals of **B** on its diagonal.

Because all dissimilarities between MP's that belong to the same party are zero, and all dissimilarities between the MP's of two different parties are equal, many ties exist in $\mathrm{D}(\mathbf{B}\mathbf{M}^{-1/2}) = \{\chi^2_{ij}(\mathbf{B})\}^{1/2}$. Therefore, the dissimilarities were transformed monotonically with Kruskal's primary approach to ties that allows ties in the data to become untied in the transformation. So the restriction is that within-party pseudo-distances remain smaller than between-parties pseudo-distances, and that the latter are monotonic with the original chi-square distances. By contrast, the 8 political issue variables were treated in a similar way as the sympathy scales in the previous application: not the dissimilarity variables but the given variables were optimally transformed to give distances $\mathrm{D}(\mathbf{q}_m)$, using second-degree monotonic splines with two interior knots.

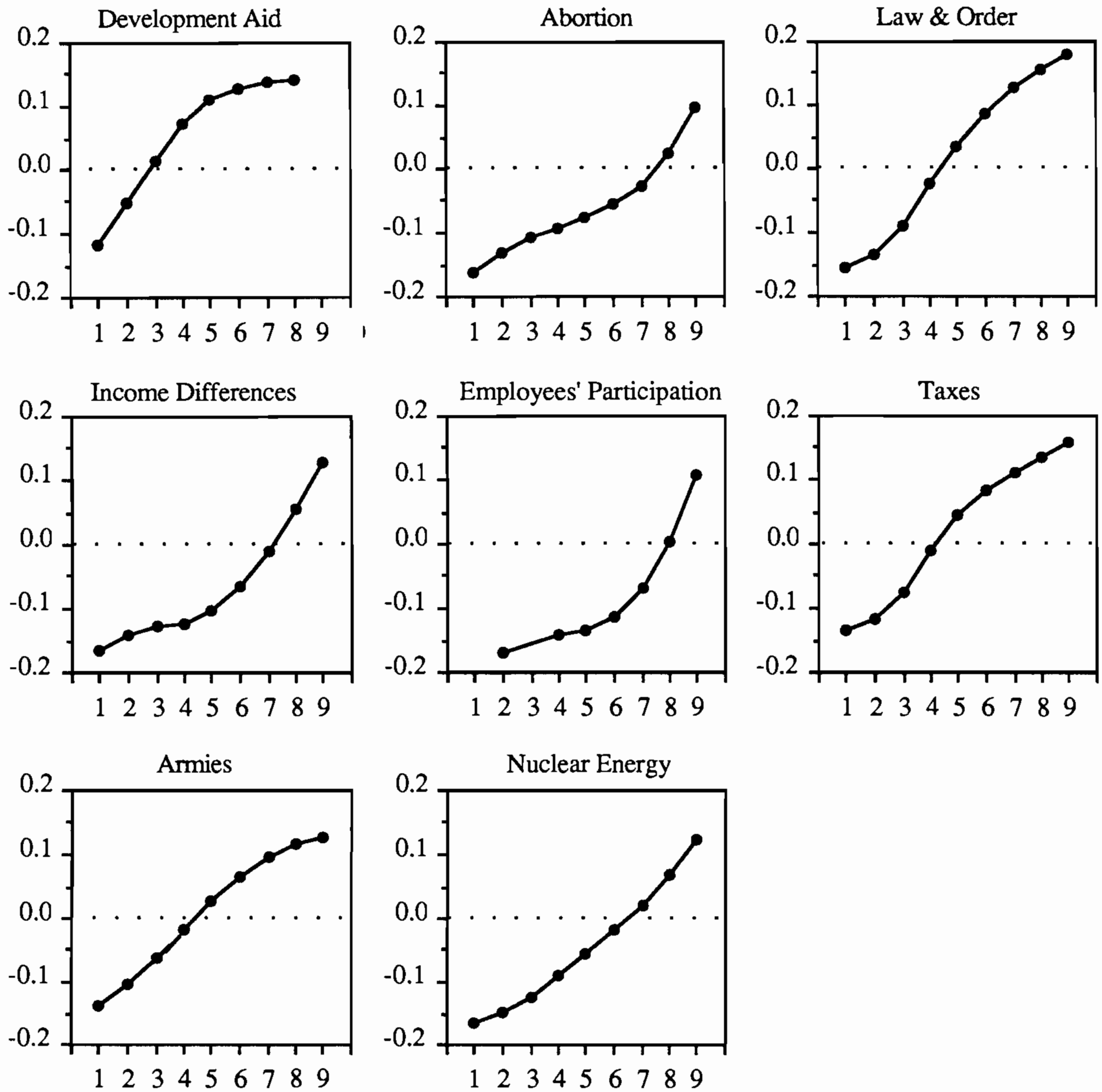

Figure 3.
Monotonic spline transformations for 8 political issues.

Figure 3 gives the transformation of the issue variables. It is important to scrutinize them, because the values on the original scales were equally-spaced, but they will in general no longer be through the monotonic spline transformations. When the transformation is a concave function as for Development Aid, the lower end of the scale (much more money) is emphasized, while the upper end (less money) is de-emphasized. When the transformation is a convex function as for Employees' Participation, the lower end (only management should decide) is de-emphasized; the upper end (employees should have their say too) is emphasized, probably because this is not a very extreme statement. When the transformations are viewed together with the marginal frequencies of the categories in Table 2, we see that the curves are steep when the associated marginal frequencies are large, while the curves are flat when the marginal frequencies are small (compare the concave function for Development Aid, with marginal frequencies 29, 35, 25 for the categories 1, 2,3 and marginal frequencies 2, 3, 1 for the categories 6, 7, 8); in short, it turns out that the optimal spline transformations follow the cumulative frequency distributions very closely.

TABLE 5
Weights and Loss for Each Source, and
Partitioned Loss: Total Stress = Heterogeneity + Group Stress

Source	First Point of View Weight	Second Point of View Weight	Loss in Applicable Point of View Stress	Heterogeneity
Law & Order	0.900		0.190	0.144
Income Differences	0.893		0.203	0.149
Employees' Participation	0.829		0.313	0.202
Taxes	0.902		0.187	0.148
Armies	0.896		0.197	0.145
Nuclear Energy	0.893		0.202	0.147
Development aid		0.895	0.199	0.164
Abortion		0.901	0.189	0.158
Party Membership		0.933	0.129	0.068
	Total Stress	Hetero-geneity	Group Stress	Mean Weight
First point of view	0.144	0.104	0.040	0.886
Second point of view	0.057	0.043	0.014	0.907
Overall	0.201	0.147	0.054	0.894

Two points of view arise from the analysis: the first is formed by the variables Law & Order, Income differences, Employees' Participation, Taxes, Armies and Nuclear Energy, and the second by Development Aid and Abortion; the monotonically transformed dissimilarities that were derived from party membership go along with the second point of view. Table 5 gives the weights, the badness-of-fit values for each source, and the partitioned loss.

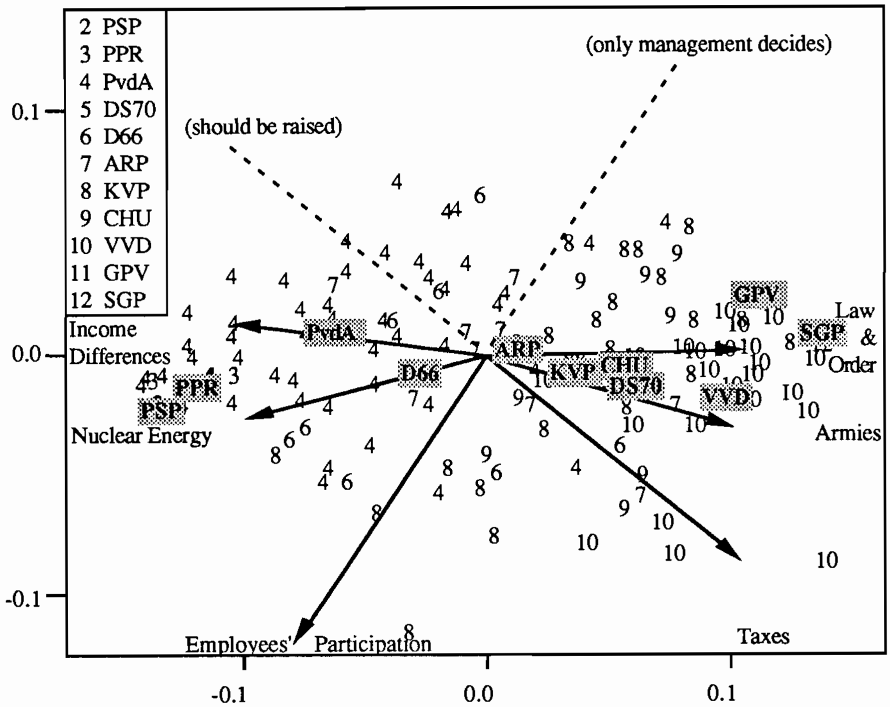

Figure 4.
Analysis of political issues and party membership:
Multidimensional structure according to Employees' Participation and Taxes
versus Income Differences, Nuclear Energy, Law & Order, and Armies

The first point of view is represented in Figure 4; the 139 MP's are represented by points with the labels from Table 1 that indicate their party membership. Next, points are given for the parties: these are the centroids of individual MP's who belong to the same party. Finally, the political issues that constitute the first point of view are represented as vectors, whose coordinates have been obtained through multiple regression, using the coordinates of the MP's as independent variables and the optimally transformed issue variables as dependent variables.

The space is in principal axes position; since the eigenvalues are 0.85 and 0.15, there is a very dominant first dimension. When the order of the parties along this dimension is compared with the ordering from left to right in Table 1, it might be considered closely related to the left-right continuum. (As in the first analysis, DS70 is positioned on the conservative side.) MP's that are positioned left from the origin are more likely to feel that income differences should be smaller, and nuclear power plants should not be built when compared to MP's right from the origin; also, they feel stronger that armies should be reduced, and that the government takes too rigorous action against disturbances of the peace. Employees' Participation and Taxes do not go along with the first dimension; various MP's of the VVD, for example, feel stronger about Employees' Participation than a considerable number of MP's of the PvdA.

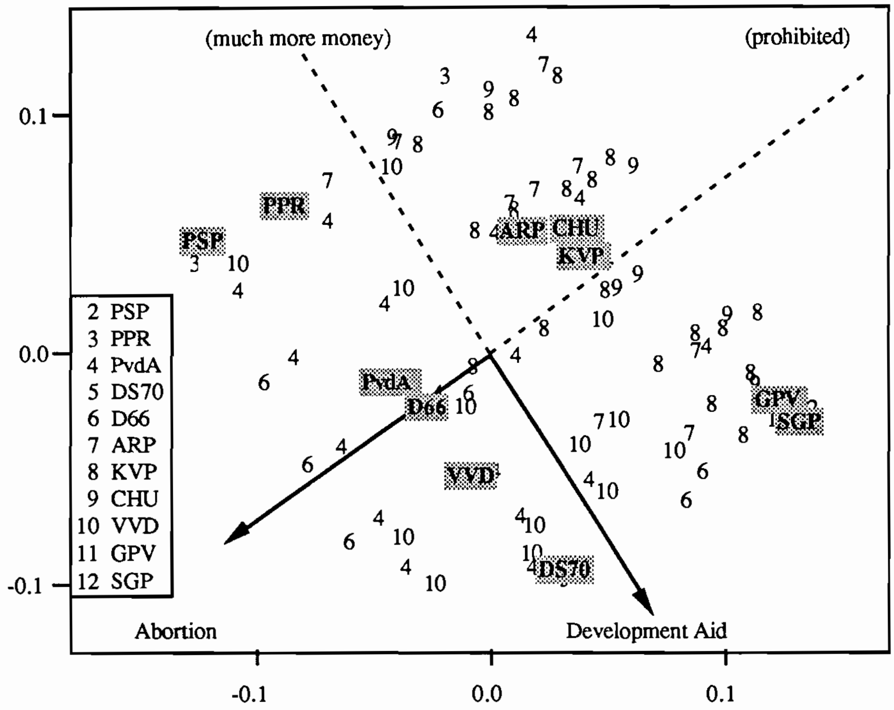

Figure 5.
Analysis of political issues and party membership:
Multidimensional structure according to Abortion, Development Aid,
and Party Membership

Figure 5 gives the position of the MP's according to Abortion and Development Aid; the configuration is certainly not one-dimensional (the eigenvalues are 0.52 and 0.48). Abortion separates the denominational parties GPV, SGP, CHU, KVP, and ARP (that feel that abortion should be prohibited) from the parties that feel that every woman has the right to decide for herself: these parties are the left-wing PSP, PPR, PvdA, and D66, but also the economically conservative VVD and DS70. The extreme positions towards Development Aid are taken by the PPR and PSP (much more money), and DS70, SGP, and GPV (less money); hardly no distinction is found between the ARP, CHU, KVP, PvdA, and D66. Inspecting the position of the party points along the first dimension, we see that the position of the (conservative) VVD is considerably different from its position in Figure 4.

As is clear from Figure 5, there is quite some heterogeneity within parties; this heterogeneity can be inspected more closely through the transformed dissimilarities according to party membership. Because ties were allowed to be untied, within-party pseudo-distances will differ from the original zero dissimilarities, and large discrepancies from zero indicate large heterogeneity. The within-party pseudo-distances were grouped into seven classes, and the frequencies for each class computed. The cumulative frequency distributions (expressed in proportions) are displayed in Figure 6 (the small parties were omitted from this graph). Compared to KVP and CHU, PvdA and VVD are more homogeneous and ARP is less so, while D66 displays a remarkable heterogeneity.

Although the party membership variable has been assigned to the second point of view, a comparable graph (Figure 7) has been constructed for the first point of view, using pseudo-distances computed for fixed weights and the configuration for the first point of view. As could be expected, the overall heterogeneity is larger than in the second point of view; VVD, ARP, and CHU are more homogeneous than PvdA and KVP, while D66 is again the most heterogeneous.

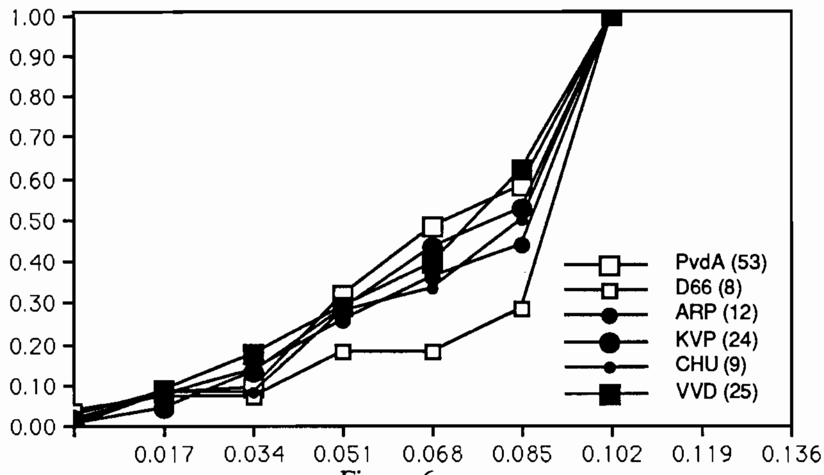

Figure 6.
Cumulative frequency distributions displaying heterogeneity within parties, derived from grouped within-party pseudo-distances.
Proportions (vertical axis) versus pseudo-distances (horizontal axis) for the second point of view.

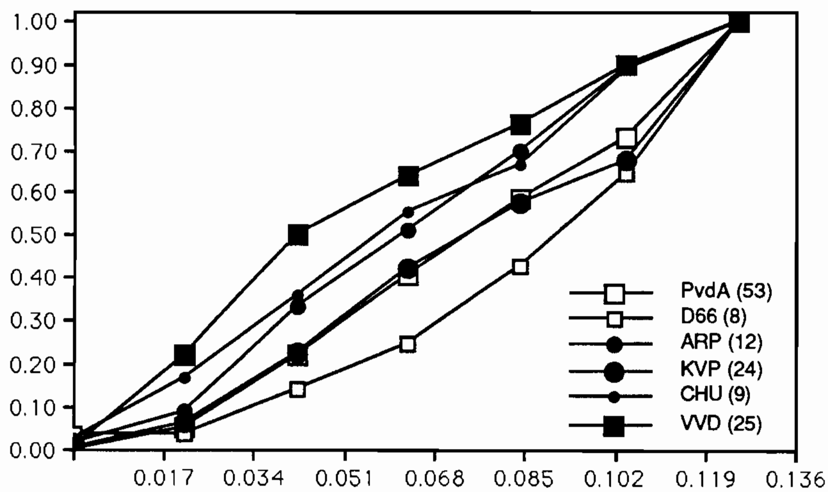

Figure 7.
Cumulative frequency distributions displaying heterogeneity within parties, derived from grouped within-party pseudo-distances.
Proportions (vertical axis) versus pseudo-distances (horizontal axis) for the first point of view.

Are Multiple Points of View Really Different?

When two or more different points of view have been obtained, their possible similarity can be investigated by the use of quadratic assignment procedures (see Hubert, 1987, for an extensive review). In the analysis of the sympathy scales, the similarity between the two spaces is captured in the correlation between the two distance vectors $\mathbf{d}(\mathbf{X}_1)$ and $\mathbf{d}(\mathbf{X}_2)$, which turns out to have a very small value of 0.042. To estimate the probability of a correlation as large or larger than 0.042 occurring by a random displacement of the points, the coordinates in $\mathbf{X}_2$ were permuted, with the number of random permutations set to 5000. The Monte Carlo distribution is given in Table 6; the p-value is 0.314, thus there is no evidence of communality between the two viewpoints.

TABLE 6
Monte Carlo Distribution of Grouped Correlations
Between Distances in First Permutation Study
Sample size of 5000, Observed Correlation is 0.042;
p-value = 0.314

Class			f	cf	f/N	cf/N
.641	-	.739	5	5	0.001	0.001
.491	-	.641	22	27	0.004	0.005
.342	-	.491	150	177	0.030	0.035
.192	-	.342	408	585	0.082	0.117
.042	-	.192	987	1572	0.197	0.314
–.108	-	.042	2194	3766	0.439	0.753
–.258	-	–.108	1230	4996	0.246	0.999
–.290	-	–.258	4	5000	0.001	1.000

It was chosen to define the similarity between the two political issues spaces on the similarity between the party points (the centroids of the MP's belonging to the same party). Because the marginal frequencies (the weights assigned to the centroids) are very different, the correlation has been computed taking these different frequencies into account. When the centroids are given in $\mathbf{Y}_1 = \mathbf{M}^{-1}\mathbf{B}'\mathbf{X}_1$ and $\mathbf{Y}_2 = \mathbf{M}^{-1}\mathbf{B}'\mathbf{X}_2$, with $\mathbf{B}$ and $\mathbf{M}$ as defined in (9), then the correlation is considered between $\mathbf{d}(\mathbf{BY}_1)$ and $\mathbf{d}(\mathbf{BY}_2)$. The observed correlation is 0.287; the Monte Carlo distribution of the grouped correlations is given in Table 7; the p-value is

0.109, so there is no strong evidence that the two sets of centroids are giving the same structure.

TABLE 7
Monte Carlo Distribution of Grouped Correlations Between Distances in Second Permutation Study
Sample size of 5000, Observed Correlation is 0.287;
p-value = 0.109

Class			f	cf	f/N	cf/N
.496	-	.683	79	79	0.016	0.016
.287	-	.496	466	545	0.093	0.109
.078	-	.287	1154	1699	0.231	0.340
−.131	-	.078	1834	3533	0.367	0.707
−.339	-	−.131	1386	4919	0.277	0.984
−.473	-	−.339	81	5000	0.016	1.000

Application of the INDSCAL Model to the Sympathy Scales.

It seems that a basic difference between the PVA model and the INDSCAL model is that INDSCAL fits individual spaces, while the PVA model fits group spaces. Of course, the common space in the INDSCAL model, rescaled by using the average weight matrix, could also be interpreted as giving dimensions of different points of view, but it might not be clear how to combine dimensions into a point of view. This will be illustrated by reanalyzing the sympathy scales with the INDSCAL model by minimizing the loss function

$$\mathrm{STRESS}(\mathbf{X};\mathbf{A}_1,\ldots,\mathbf{A}_M) = M^{-1} \sum_{m=1}^{M} \|\mathbf{\Delta}_m^* - \mathrm{D}(\mathbf{X}\mathbf{A}_m)\|^2,$$

over the common space $\mathbf{X}$ and the diagonal weight matrices $\mathbf{A}_1,\ldots,\mathbf{A}_M$ for given $\mathbf{\Delta}_m^* = \mathrm{D}(\mathbf{q}_m)$, with the latter obtained from the points of view analysis. A special purpose algorithm was developed, based on the majorization approach detailed in Heiser and Stoop (1986). There is some freedom to choose from different, but coherent, normalizations; because we wish the common space to have an explicit shape, the common space was not normalized but instead the weights were normalized so that $M^{-1} \sum_m A_m^2 = \mathbf{I}$ (this is equivalent to non-normalized weights and normalized dimensions with sum of squares of one).

Since there were 2x2 dimensions in the points of view analysis, the INDSCAL model was fitted with 4 common dimensions; the stress in the INDSCAL model is 0.145, so the 139 individual spaces (each using four parameters) fit the data only slightly better than the 2 point of view spaces (using 1 parameter per source, with the total stress 0.183). The INDSCAL stress equals 1–0.855, the sum of squares of the coordinates across dimensions (see Table 8).

TABLE 8
Correlations between INDSCAL and PVA dimensions

		Dimensions First Point of View		Dimensions Second Point of View		Rotated First Point of View		Rotated Second Point of View	
INDSCAL	SSQ	1	2	1	2	1	2	1	2
1	0.301	–0.01	–0.71	0.98	–0.15	–0.13	–0.31	0.99	0.61
2	0.285	0.98	0.12	–0.07	–0.26	0.99	–0.83	–0.03	–0.23
3	0.136	–0.85	0.41	–0.21	0.61	–0.77	0.95	–0.32	0.26
4	0.133	0.16	0.04	0.66	0.63	0.16	–0.12	0.54	0.91

The correlations between the dimensions of the INDSCAL common space and the 2x2 dimensions in the point of view spaces are given in columns 2–5 of Table 8. It is clear that points of view analysis and INDSCAL have two dimensions in common; it is unclear what happened to the other two point of view dimensions. Since INDSCAL is in its unique orientation of axes, and the points of view are in principal axes orientation, correlations between dimensions may be quite meaningless. Computing the correlations between the distances in the two points of view and the distances derived from all possible pairs of dimensions in the INDSCAL common space, shows that the INDSCAL distances derived from dimensions 2 and 3 correlate 0.98 with the distances in the first point of view, and that those in dimensions 1 and 4 correlate 0.97 with the distances in the second point of view. Rotating the dimensions of the PVA spaces to these two pairs of INDSCAL dimensions gives the correlations between dimensions given in the columns 6–9 of Table 8. Now it is clear that points of view and INDSCAL do find the same structure.

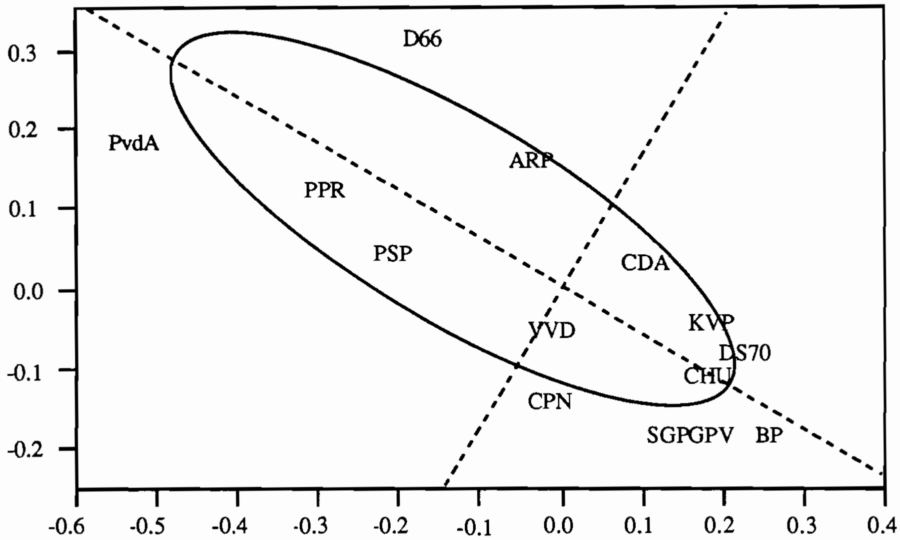

Figure 8. INDSCAL analysis of political sympathy scales: dimension 3 (vertical axis) versus dimension 2 (horizontal axis). Dotted lines and ellipse indicate the orientation of the left-wing point of view.

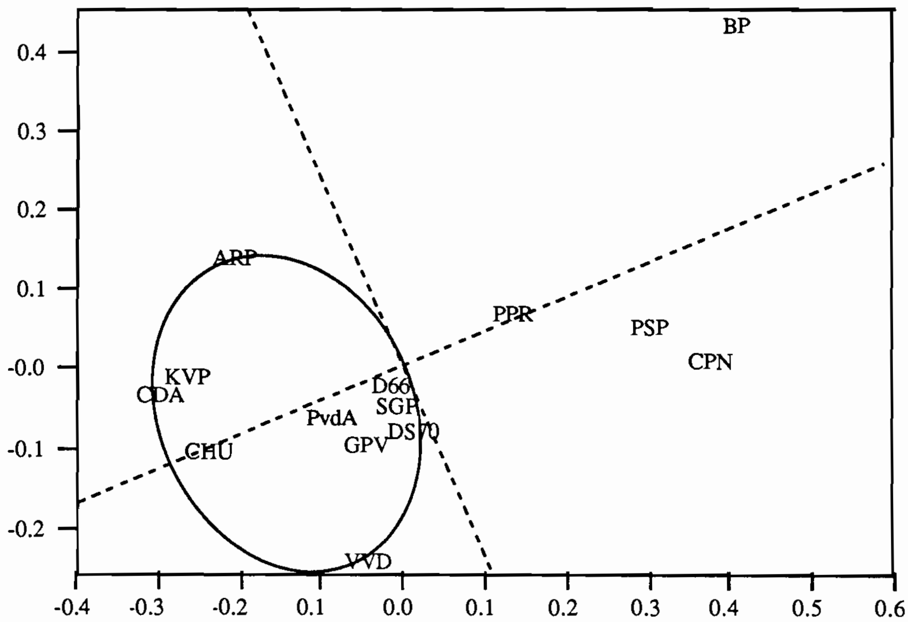

Figure 9. INDSCAL analysis of political sympathy scales: dimension 4 (vertical axis) versus dimension 1 (horizontal axis). Dotted lines and ellipse indicate the orientation of the center-right-wing point of view.

The dimensions of the INDSCAL common space are displayed in Figure 8 (dimension 3 versus dimension 2) and Figure 9 (dimension 4 versus dimension 1). To compare the solution with the point of view spaces, the elliptical structures drawn in the Figures 1 and 2 were rotated to display the position of the axes of the point of view spaces. There is a strong similarity, although not a perfect match; consider, for example, the different positions of PvdA and D66 in the Figures 2 and 9. It is unclear, however, why INDSCAL uses its particular orientation of the axes; there does not seem to be a substantive interpretation of the dimensions of the common space.

5. Discussion

The primary purpose of this paper has been to show that the concept of points of view analysis is worthwhile for the analysis of heterogeneous sources on a group level; since sources may be homogeneous within groups and heterogeneous between groups, the strength of the points of view analysis concept is in its parsimonious display of objects in $r \geq 2$ points of view, when it is possible to group sources into subsets because they share a particular viewpoint about the objects' interrelationships.

Tucker and Messick's original procedure was called an individual differences model; we agree with Carroll and Chang (1970) that if the objective of analysis is individual differences scaling, application of the INDSCAL model would be more appropriate, since individual differences are displayed in separate spaces and each individual source has the possibility to be distinct from all other sources. Although this is clearly a subject of further study, the INDSCAL common space may not be the best way to display differences at the level of groups, since the INDSCAL model, as a truly individual difference model, allows each individual to weight each dimension of the common space differently. More restricted INDSCAL models have been proposed, requiring, for example, that each individual source may use only $t < p$ dimensions, where p denotes the dimensionality of the common space. However, this restriction does not guarantee that sources in an intrinsically homogeneous group will use exactly the same t dimensions, because dimensions in the INDSCAL model are allowed to be correlated. Our conjecture would be that further restrictions on the INDSCAL model to find homogeneous groups would lead to the PVA model.

Kiers (1989) discusses the relationship between various approaches to three-way scaling from a different perspective. He notes there is resemblance in lay-out between Tucker and Messick's original approach and the French method STATIS (based on Escoufier, 1973), but finds them clearly different in several respects. From our perspective, the similarity is obvious. The basic idea is the treatment of (dis)similarity matrices as variables from which linear combinations are formed and subsequently subjected to a secondary analysis. In STATIS the problem remains how to avoid negative weights in a subsequent linear combination when the first composite matrix has been taken out. Another approach is found in Escoufier (1988), who proposes to find non-overlapping subsets of variables to obtain different composites, similar to the constraints placed on the weights in our procedure.

There is also a relationship with what is called 'homogeneity analysis as a first step' in Gifi (1990, ch. 3), where r different quantifications of the categorical variables are first obtained that are subsequently used to obtain r principal components analyses solutions in p_s dimensions. Related work on quantifying categorical variables in a three-way framework is Saporta (1975), and the extensive review in Kiers (1989).

The procedure described in this paper could easily be generalized to allow sources to be assigned to t points of view, where $1 \leq t < r$ (giving overlapping clusters). When variables are optimally transformed and assigned to more than one point of view, one could choose either identical or possibly different optimal transformations with respect to the different points of view. This extension, however, needs further investigation with respect to its data analytical merits. Another possible extension would allow the given matrices $\mathbf{\Delta}_m$ to be asymmetric.

In the procedure described, the (dissimilarity) variables were assumed to constitute different points of view, and differential aggregation over homogeneous sources was applied to obtain different composite matrices optimal with respect to distances. It is important to realize that this basic idea of differential aggregation is very general, and can be applied to any other multivariate analysis technique, as was indicated for the principal components analysis in (6). Also, three-way generalizations to sets of variables are feasible; Verboon and Heiser (1990) have applied a similar idea to the analysis of dynamic three-way data.

The present technique gives overall dissimilarity measures in $\mathbf{\Theta}_s$, as an aggregation over a group of homogeneous sources. When the sources $\mathbf{\Delta}_m^*$ are homogeneous at the outset, we choose s equal to 1. Related work in finding such a composite matrix is given by Escoufier (1980) and Gower (1971). A related procedure of differentially weighting variables to find an optimal representation is found in De Soete, DeSarbo, and Carroll (1985), who simultaneously estimate variable importance weights and the corresponding ultrametric tree.

The procedure proposed in the present paper combines aspects of the work just referenced; it finds clusters of sources, and weights the sources differentially to find composite matrices. The individual sources may be given directly or may be derived from numerical variables, ordinal variables (for which a monotonic transformation is obtained) or nominal variables, either by obtaining nominal transformations or by using indicator matrices. Although the composite matrix $\mathbf{\Theta}_s$ is optimal for least squares distance fitting, it could also be analyzed afterwards by other techniques, for example by a cluster analysis.

Acknowledgement

This research was supported by the Royal Netherlands Academy of Arts and Sciences (KNAW). An earlier version of this paper was presented at the European Meeting of the Psychometric Society, Leuven, 1989. We wish to thank Willem J. Heiser for his stimulating comments.

References

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283-319.

Cliff, N. (1968). The "idealized individual" interpretation of individual differences in multidimensional scaling. *Psychometrika, 33*, 225-232.

De Leeuw, J. (1988). Convergence of the majorization algorithm for multidimensional scaling. *Journal of Classification, 5,* 163-180.

De Leeuw, J., & Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In P.R. Krisnaiah (Ed.), *Multivariate analysis, Vol. V* (pp. 501-522). Amsterdam: North-Holland.

De Soete, G., DeSarbo, W.S., & Carroll, J.D. (1985). Optimal variable weighting for hierarchical clustering: an alternating least-squares algorithm. *Journal of Classification, 2,* 173-192.

Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics, 29,* 751-760.

Escoufier, Y. (1980). Exploratory data analysis when data are matrices. In K. Matusita (Ed.), *Recent developments in statistical inference and data analysis* (pp. 45-53). Amsterdam: North-Holland Publishing Company.

Escoufier, Y. (1988). Beyond correspondence analysis. In H.H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 505-514). Amsterdam: North-Holland Publishing Company.

Gifi, A. (1990). *Nonlinear multivariate analysis* [first edition 1981, Dept. of Data Theory, The University of Leiden]. Chichester: Wiley.

Gower, J.C. (1971). A general coefficient of similarity and some of its properties, *Biometrics, 27,* 857-871.

Heiser, W.J., & Stoop, I. (1986). *Explicit SMACOF algorithms for individual differences scaling*. Research Report RR-86-14. Leiden: Dept. of Data Theory.

Hubert, L.J. (1987). *Assignment methods in combinatorial data analysis*. New York: Marcel Dekker.

Kiers, H.A.L. (1989). *Three-way methods for the analysis of qualitative and quantitative two-way data*. Leiden: DSWO Press.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1-28.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika, 29*, 115-129.

Kruskal, J.B., Young, F.W., & Seery, J.B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding*. Murray Hill: Bell Laboratories.

Meulman, J.J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.

Ramsay, J.O. (1982a). Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society, Series A, 145,* 285-312.

Ramsay, J.O. (1982b). *MULTISCALE II Manual*. Dept. of Psychology, McGill University, Montréal.

Ramsay, J.O. (1989). Monotone regression splines in action. *Statistical Science, 4,* 425-441.

Ross, J. (1966). A remark on Tucker and Messick's "Points of view" analysis. *Psychometrika, 31*, 27-32.

Saporta, G. (1975). Pondération optimale de variables qualitatives en analyse des données [Optimal weighting of qualitative variables in 'analyse des données']. *Statistique et Analyse des Données, 3*, 19-31.

Shepard, R.N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function I. *Psychometrika, 27*, 125-140.

Shepard, R.N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function II. *Psychometrika, 27*, 219-246.

Tucker, L.R (1951). *A method of synthesis of factor analysis studies*. Personnel Research Section Report No. 984. Washington, D.C.: Department of the Army.

Tucker, L.R, & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika, 28,* 333-367.

Verboon, P., & Van der Kloot, W.A. (1989). *On the relation between INDSCAL and nonlinear points of view analysis*. Research Report PRM-89-02. Leiden: Department of Psychometrics.

Verboon, P., & Heiser, W.J. (1990). *Some possibilities for the analysis of dynamic three-way data*. Research Report RR-90-01. Leiden: Department of Data Theory.

Winsberg, S., & Ramsay, J.O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika, 48*, 575-595.